JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

QUANTIFICATION UNCERTAINTY IN PLASTIC CREDIT SYSTEMS: A REPRODUCIBLE ASSESSMENT OF MEASUREMENT CONTROLS, MATERIAL RECONCILIATION AND ELIGIBLE QUANTITIES

Chinmaya Bhat
Independent Researcher
Plastic Credit Systems and Sustainable Waste Management, Karnataka, India
chinmayapaaras@gmail.com

Abstract : A plastic credit is denominated in mass, so every credit rests on a weighing operation. This study examines how four programmes govern that operation and what their differences are worth. The measurement, calibration, contamination and reconciliation requirements of the Verra Plastic Waste Reduction Standard, Ocean Bound Plastic (OBP), the PCX Solutions Plastic Pollution Reduction Standard and the Circular Credits Mechanism (CCM) were compared at clause level, and the eight quantification equations of Verra's methodologies PWRM0001 and PWRM0002 were transcribed after text extraction corrupted every mathematical glyph, six of the eight by rendering the source page as an image and the remaining two from extracted text that was clean for those two equations. Measurement error was then propagated through those equations by Monte Carlo simulation under each programme's actual configuration of controls. Two of the four programmes, OBP and CCM, were found to contain no weighing-equipment calibration provision of any kind; three contain no moisture or contamination deduction. The simulation shows that the two principal controls act almost independently: a dry-weight rule reduces mean bias in credited mass from 0.85 per cent to 0.22 per cent while barely affecting spread, whereas the calibration interval leaves mean bias unchanged — the paired contrast is within its own Monte Carlo error — and widens the standard deviation from 0.274 to 3.211 per cent, the lower figure being a zero-drift benchmark rather than an achievable practice. Configured as each programme specifies, expected bias is 0.23 per cent for Verra, 0.86 per cent for PCX and 0.83 per cent for OBP and CCM, whose 95 per cent interval reaches 6.95 per cent; the PCX and OBP means are statistically indistinguishable and what separates them is dispersion. A separate finding concerns Verra's own internal asymmetry: PWRM0001 sets a capacity-addition baseline on the maximum of three prior years while PWRM0002 uses the average, and simulation shows the average convention yields a credited net 41 per cent higher at the mean. The study sets out in full the simulation protocol and a measurement-control checklist.

IndexTerms - quantification uncertainty, measurement error, calibration, dry weight, mass balance, plastic credits, Monte Carlo simulation, baseline

I. INTRODUCTION

Plastic credits are sold by the tonne. Whatever else a programme certifies, the number on the credit is a mass, and that mass originates in a physical weighing operation performed at a collection point, a weighbridge or a processing line. Everything downstream, the registry entry, the retirement record and the corporate claim, inherits whatever error that operation contained.

Mass is measured repeatedly along a recovery chain, and the measurements do not agree. Material is weighed at collection, again on receipt, again after sorting, again at dispatch and again at processing. Between those points it loses moisture, sheds non-plastic debris, has rejects removed and incurs process losses. A difference between two weighings is therefore not necessarily an error; the question is whether it is explained. Scale calibration drift, stale tare weights, inconsistent units and estimates substituted for missing measurements are errors, and they are not distinguishable from legitimate process loss without controls designed to separate them.

These errors do not all behave the same way, and the distinction matters for control design. A scale that drifts in an unknown direction adds variance: individual measurements become less reliable, but over many consignments the errors tend to cancel. Weighing wet material as though it were dry adds bias: every affected measurement errs in the same direction, and the errors accumulate rather than cancel. A control that addresses one does not address the other. Which of the two a programme regulates, and whether it regulates both, is therefore a substantive design question rather than a matter of administrative detail.

The methodological problem for research is the familiar one. Weighbridge tickets, production logs and inventory records from operating facilities are commercially sensitive and are not published, so the actual distribution of measurement error in plastic

recovery cannot be observed from outside. What can be done, and what this study does, is to take the quantification equations the programmes publish, state a set of error assumptions explicitly, propagate those assumptions through the published equations, and report how the result changes as the control regime changes. The output is a statement about control logic under stated assumptions, not a measurement of industry error.

One practical obstacle had to be overcome first. Text extraction from both Verra methodologies corrupted every mathematical-italic glyph, rendering variable names as mojibake and separating summation indices from their operators. Equations recovered from such text cannot be trusted, and an earlier reading inferred the structure of two of them from surrounding prose rather than reading it. Six of the eight equations were therefore confirmed by rendering their source pages as images and reading the typeset equations directly, and the status of each is recorded individually with the page it was read from. The two exceptions are the baseline and project collection equations of PWRM0001, which are plain single summations with no stacked fraction, no adjustment factor and no nested bracket, and whose extracted text was legible; these are reported as text transcriptions rather than as verified ones, and a reader relying on their exact layout should check them against the published page. Two findings emerged from the rendering exercise which would not have survived a text-only reading, and both are reported below.

The study contributes a clause-level comparison of measurement and reconciliation requirements; a transcription of the eight quantification equations with their defaults; a taxonomy separating measurement, data-recording and methodological uncertainty; a Monte Carlo propagation showing what each control is worth; a quantification of an internal asymmetry between Verra's two methodologies; and a measurement-control checklist.

II. MEASUREMENT CONTROLS ACROSS THE FOUR PROGRAMMES

Table 1: Measurement and reconciliation controls compared at clause level

Measurement control	Verra	OBP	PCX	CCM
Weighing-equipment calibration required	Yes, manufacturer spec or at least every three years	No provision located	Yes, calibrated equipment mandatory	No provision located
Calibration evidence submitted with claim	Not specified	Not applicable	Yes, proof uploaded at issuance	Not applicable
Dry-weight or contamination deduction	Yes, dry weight net of non-plastic substances	No numeric limit; best-efforts only	No deduction located	No correction located
Quantification equation published	Yes, eight equations	Mass-balance rules only	None located	None beyond 1 tonne = 1 credit
Numeric uncertainty discount or conservativeness factor	None	None	None	None
Mass-balance reconciliation tolerance	None stated	Mass balance required, no tolerance	Discrepancy analysed, no tolerance	No reconciliation equation stated
Credit denominated in	One tonne	One kilogram (DEF-GUI s.5)	One tonne	One tonne
Default values supplied where data absent	Yes, waste generation rates and composite fractions	No	No	No

Table 1 compares the measurement controls. The pattern is not one of varying stringency but of different controls being present and absent in different combinations.

Calibration divides the four sharply. Verra [1], [2] requires that weighbridges be “calibrated according to the equipment manufacturer's specifications or at least every three years”, with a general obligation that monitoring equipment be maintained and calibrated according to current good practice. PCX [7] goes further in one specific respect: calibrated equipment is mandatory, its Module 4 record field specifies that the collection record should reflect the “identifier of the weighing equipment” [7], which is worded as guidance rather than obligation, and “the proof of calibration for the weighing equipment shall be presented and uploaded into the PPRS Registry” at issuance. Requiring the evidence at the moment of claim, rather than requiring the practice and leaving verification to audit, is the strongest calibration control located in any of the four programmes.

Against that, no provision requiring weighing equipment to be calibrated, requiring a calibration certificate, or establishing metrological traceability was located anywhere in the Ocean Bound Plastic documents [4], [5], [6] or in the Circular Credits Mechanism documents [8], [9]. Two OBP provisions come near the subject without supplying such a rule: the audit protocol refers to a tolerance in the context of reconciling audited quantities, and the collection standard contemplates recording an “approximate weight per bag”, which presupposes weighing but regulates neither the instrument nor its accuracy. This is an absence claim and is stated with the search behind it: it rests on full reading of ten OBP instruments and six CCM instruments. The nearest provisions are descriptive equipment-disclosure fields in CCM's validation checklist and monitoring template, which ask the project to “specify type of equipment, owner of equipment or source of data” without imposing any requirement on the equipment's accuracy.

Moisture and contamination divide them differently, and here Verra stands alone. PWRM0001 [2] requires that “the dry weight of the final output of collected plastic waste that reaches an appropriate end destination is directly measured and recorded”, and defines dry weight as “the difference between the weight of the collected plastic waste and the weight of any non-plastic substances”, with examples including mud, seaweed and sand. No moisture, residue or contamination deduction was located in the OBP, PCX or CCM corpora; OBP's treatment of contamination is effort-based, requiring best efforts without a numeric limit.

Conservativeness is addressed in principle by two of the four programmes and quantified by none. Verra reaches it in three places, which together go wider than the baseline. Its Plastic Standard lists conservativeness among the principles at clause 2.2 of that standard — “use conservative assumptions, values and procedures to ensure that net plastic waste collection and recycling are not overestimated” — and clause 3.1 makes those principles binding, providing that projects “shall be guided by the principles set out in that section”. Clause 3.9 then states it as a rule of baseline construction specifically: “in developing the baseline scenario, assumptions, values and procedures shall be selected that help ensure that net plastic waste collected and/or recycled is not

overestimated". Clause 3.15 adds that methodology deviations "shall not negatively impact the conservativeness of the quantification" except where they increase accuracy. The principle therefore governs quantification generally and not only the baseline. What it does not do is set a number: it directs how a judgement is to be exercised, and the same standard notes that accuracy is "difficult to attain in many cases" and that conservativeness "may serve as a moderator" of it. Separately, PWRM0002 requires that "in case of inconsistencies, a conservative approach to correction must be taken", at clause 9.3 of that methodology; no equivalent sentence appears in PWRM0001, so the rule applies to recycling projects and not to collection projects. PCX names Conservativeness as a guiding principle but words it as guidance rather than obligation: claims "should be as accurate as possible" and, "in cases where there is doubt, organizations agree to undertake the more conservative approach". What no programme does is attach a numeric deduction, discount factor or buffer to the credited quantity, and none states a numeric tolerance for mass-balance reconciliation. PCX [7] requires that "any discrepancy within the chain of custody shall be duly analyzed and explained" but sets no permitted variance, so the threshold at which a discrepancy becomes reportable is left to the verifier. CCM's monitoring template [10] carries three tables for input, output and losses, each with a total row, but states no relationship between the three totals and gives no reconciliation equation.

III. THE PUBLISHED QUANTIFICATION EQUATIONS

Verra is the only programme in the corpus that publishes a full set of quantification equations, though it is not the only one with a formula: OBP states a mandatory chemical-recycling rule that "the resulting Claimed OBP quantity must be equal to the OBP input less the process loss", which is a deduction no Verra equation contains. The credited quantity in both Verra methodologies is a simple difference between project and baseline with no further adjustment.

All eight equations published across the two methodologies are reproduced below, four from each. How each was read matters, because text extraction from both PDFs corrupted every mathematical-italic glyph, so no equation could be taken from extracted text alone. Six of the eight were therefore confirmed by rendering the source page as an image at 150 to 160 dots per inch and reading the typeset equation directly; these are marked verified with the page on which they appear. The remaining two — PWRM0001 Equations 2 and 3 — are simple summations whose surrounding definition lists extracted cleanly, and they are marked as read from text. Both penetration-rate equations were verified from rendered pages, which matters because the cross-methodology difference reported below turns on them.

For collection, PWRM0001 [2] first states the penetration-rate formula used at Step 3a of its additionality procedure, then aggregates baseline and project quantities by material type and nets them by destination. Verra's own numbering is used throughout.

$$\text{Penetration Rate} = \left(\frac{C}{G}\right) \times 100 \qquad \text{(PWRM0001 Eq. 1)}$$

$$B_{\text{collected,d,y}} = \sum_{i=1}^{n} B_{\text{collected,i,d,y}} \qquad \text{(PWRM0001 Eq. 2)}$$

$$P_{\text{collected,d,y}} = \sum_{i=1}^{n} P_{\text{collected,i,d,y}} \qquad \text{(PWRM0001 Eq. 3)}$$

$$N_{\text{collected,y}} = \sum_{d=1}^{m} \left(P_{\text{collected,d,y}} - B_{\text{collected,d,y}}\right) \qquad \text{(PWRM0001 Eq. 4)}$$

PWRM0001 Equation 1 verified from the rendered page 19; Equations 2 and 3 read from extracted text; Equation 4 verified from the rendered page 24.

In Equation 1, C is annual plastic waste collection in the region including composite materials and G is total generation of plastic waste in the region, both in tonnes per year. In Equations 2 to 4, B and P are the baseline and project quantities of material type i transferred to destination d in year y, in tonnes. The methodology does not define the summation limits n and m in the variable lists that accompany these equations; they are read here as the number of material types and the number of appropriate end destinations respectively, which is the only reading consistent with the indices, but the reading is the author's and not the document's. That the baseline term carries the destination index d establishes that it falls inside the summation over destinations; this was confirmed from the rendered page, the extracted text having placed the subtraction on a separate line.

For recycling, PWRM0002 [3] states the same penetration-rate formula, then applies two adjustment factors that have no counterpart in the collection methodology, and then takes a single difference of annual aggregates:

$$\text{Penetration Rate} = \left(\frac{C}{G}\right) \times 100 \qquad \text{(PWRM0002 Eq. 1)}$$

$$B_{\text{recycled,y}} = \sum_{i=1}^{n} B_{\text{p,recycled,i,y}} \times AF_i + \sum_{i=1}^{n} B_{\text{dp,recycled,i,y}} \times MF \qquad \text{(PWRM0002 Eq. 2)}$$

$$P_{\text{recycled,y}} = \sum_{i=1}^{n} P_{\text{p,recycled,i,y}} \times AF_i + \sum_{i=1}^{n} P_{\text{dp,recycled,i,y}} \times MF \qquad \text{(PWRM0002 Eq. 3)}$$

$$N_{\text{recycled,y}} = P_{\text{recycled,y}} - B_{\text{recycled,y}} \qquad \text{(PWRM0002 Eq. 4)}$$

PWRM0002 Equation 1 verified from the rendered page 23, Equation 2 from page 26, and Equations 3 and 4 from page 30.

AF_i is an adjustment factor for composite material i, equal to 1 for non-composite materials, and MF is the mass fraction of the output of a depolymerization process used for plastic production. PWRM0002 Section 8.1 supplies default plastic fractions for composite materials, applied through AF_i: 0.04 for unspecified composite materials, 0.20 for used beverage cartons, 0.05 for paper cups and 0.04 for e-waste. No default is published for MF. The methodology requires projects to calculate it from the actual mass used for plastic production, and its worked example yields 0.5 for a steam cracker half of whose output goes to polymers; for depolymerized material it notes that "since the material type can no longer be determined based on the output, it must be determined based on the input to the depolymerization process, using a mass balance approach".

Two findings emerged from verifying these equations that a text-only reading would have missed, and both bear on comparability.

The first concerns the penetration-rate equation, which both methodologies state in identical form as the ratio of C to G multiplied by 100, with the same 20 per cent decision threshold and the same two Step 3a outcomes. The symbols are the same and the formula is the same, but C denotes different quantities. In PWRM0001 the numerator is annual plastic waste collection; in PWRM0002 the governing sentence reads that the rate "is the ratio between the total installed recycling capacity and total waste

generation". Collection is a flow that is realised; installed capacity is a stock that may or may not be used. A reader comparing the two methodologies on the printed formula alone would conclude they apply the same test.

The second is a notation inconsistency within PWRM0002. The left-hand side of its baseline equation is B_recycled,y, aggregated over material types, while the monitored parameter in the monitoring table at clause 9.1 of that same methodology is labelled B_recycled,i,y and described as the "amount of material type i recycled in the baseline in year y". The equation aggregates; the monitoring table does not. Both labels were confirmed by rendering their respective pages. The two are different quantities and the document uses both without reconciling them.

IV. A TAXONOMY OF QUANTIFICATION UNCERTAINTY

Table 2 separates the uncertainty affecting a credited quantity into three classes, distinguished by what produces them and therefore by what controls them.

Measurement uncertainty arises in the physical weighing operation. Most of it adds variance, because a drifting scale or a noisy reading errs in an unpredictable direction and the errors tend to cancel across many consignments. Moisture is the exception within this class: wet material and adherent debris always read high, so the error accumulates.

Data-recording uncertainty arises after the weighing, in how the value is transcribed, stored and corrected. Its characteristic form is rare and large rather than frequent and small: a kilogram figure recorded as tonnes is wrong by three orders of magnitude, which makes it both more dangerous and easier to detect than measurement noise. An undocumented correction is the irreducible case, because the recorded value cannot be verified in either direction and the uncertainty is not quantifiable at all.

Methodological uncertainty arises from the rules themselves rather than from any error in applying them. A project measured perfectly and recorded perfectly will still yield different credited quantities under different baseline conventions, and the difference is attributable to the methodology rather than to the facility. Section 6.4 quantifies one instance of this.

Table 2: Classes of uncertainty affecting a credited quantity, their behaviour and the control that addresses each

Class	Source of uncertainty	Behaviour	Control that addresses it
Measurement	Scale calibration drift	Bias within one instrument; variance across instruments	Calibration interval and proof
Measurement	Weighing noise and resolution	Variance	Equipment specification
Measurement	Stale or wrong tare weight	Bias for as long as the tare stands	Tare re-measurement procedure
Measurement	Moisture and non-plastic debris	Bias, always upward	Dry-weight rule and deduction
Data recording	Unit transcription error	Rare, very large, one-directional	Unit plausibility check
Data recording	Missing measurement replaced by estimate	Bias, typically upward	Measurement completeness rule
Data recording	Undocumented correction to a recorded value	Unquantifiable	Correction reference requirement
Methodological	Choice of baseline convention	Bias, direction set by the convention	Prescribed baseline method
Methodological	Composite and depolymerized adjustment factors	Bias if defaults misapplied	Material-specific defaults
Methodological	Absence of an uncertainty discount	Bias retained in full	Conservativeness factor; none present

V. RESEARCH METHODOLOGY

The simulation replicates a project 2,000 times per configuration and recomputes the whole quantification from first principles on every replication. A collection project is built as Verra builds it, by material type and end destination: six material types across three end destinations, with forty consignments weighed for the baseline and forty for the project in each combination, which is 1,440 weighings per replication. A recycling project carries the same six material types in a project stream and a baseline stream, forty consignments each, with a further ten depolymerized consignments per material type per stream, which is 600 weighings. Baseline consignments are drawn uniformly between 80 and 900 kilograms and project consignments uniformly between 250 and 2,400 kilograms, the difference in range representing the capacity addition the project is credited for; depolymerized consignments are drawn from the same range and scaled to 30 per cent. Each true mass is then subjected in sequence to six of the ten error sources in Table 2 — calibration drift, weighing noise, tare error, moisture and debris, unit transcription and a missing measurement replaced by an estimate. The remaining four are not simulated and are not meant to be: an undocumented correction is by definition unquantifiable, the baseline convention is treated separately in Section 6.4, the adjustment factors are applied as published rather than misapplied, and the absence of an uncertainty discount is what the whole exercise measures rather than an input to it. The quantities are aggregated through the published equations, and the reported net is compared against the true net to give a bias for that replication. The generator is seeded, replication r using seed 20261030 plus r, so every figure reported can be regenerated exactly.

Four constants in the recycling pathway are the author's and are set out here rather than left in the code. The adjustment factors AF_i are taken from the PWRM0002 defaults where the methodology publishes one, so used beverage cartons carry 0.20 and unspecified composites 0.04, with 1.00 for the four non-composite resin types modelled. The mass fraction MF is set to 0.85. PWRM0002 publishes no default for MF and requires projects to compute it from the actual mass used for plastic production, so 0.85 is a choice, not a provision; it was checked against the alternatives 0.50, which is the value in the methodology's own worked example, and 1.00, and the mean net bias on the recycling pathway moves from 0.735 per cent to 0.734 per cent to 0.734 per cent across the three, a range smaller than the Monte Carlo error, because MF multiplies the project and baseline depolymerized streams alike and largely cancels in the difference. The depolymerized stream is set at 30 per cent of the corresponding non-depolymerized consignment, and three end destinations are assumed for collection projects. None of these is stated by any programme.

Scale drift is modelled as accumulating with time since the last calibration. A drift rate of 0.0008 proportional bias per thirty days of uncorrected operation is applied, with time since calibration drawn uniformly across the interval the regime permits. Both

the drift direction and its magnitude are drawn once per weighing instrument and then held fixed across every consignment that instrument weighs, because a scale that reads high reads high all day. This matters: redrawing the direction at each weighing would make drift average away within a project by construction, and the finding that calibration controls dispersion rather than bias would then be an artefact of the generator rather than a property of the error. A collection project is given one weighbridge per end destination, three in all, and a recycling project one facility weighbridge. A mis-set tare is treated the same way, as a standing offset of the instrument rather than a fresh draw at each weighing. Four calibration regimes are simulated: an ideal reference with no drift; proof at issuance, with a ninety-day effective interval, matching PCX; three-yearly, matching Verra; and none, with a ten-year effective interval, standing for OBP and CCM. Only one of these four intervals is taken from a document. Verra states calibration "at least every three years" in both methodologies, which fixes the three-yearly regime at 1,095 days. PCX requires proof of calibration to be uploaded at issuance but states no interval, and the ninety-day figure is this study's reading of what an issuance-linked proof implies about the age of the last calibration; it is an assumption, not a provision. OBP and CCM state no calibration requirement at all, and the ten-year figure stands for an unregulated interval rather than for anything either programme says. The consequence is that the ordering of the four regimes follows from the documents but the size of the gaps between them does not, and the PCX result in particular would move with a different reading of what proof at issuance implies. The ideal regime is a reference point with no counterpart in any programme. Weighing noise is proportional with a standard deviation of 0.004. A stale tare weight occurs with probability 0.08 and perturbs the figure by up to 3 per cent in either direction. Material is weighed wet with probability 0.25, reading between 1 and 5 per cent high; where a dry-weight rule applies, 85 per cent of that error is removed, the residue representing imperfect drying and sorting. A unit transcription error occurs with probability 0.002 and, where a unit plausibility control is present, is caught and corrected. A measurement is missing with probability 0.02 and is replaced by an estimate that runs 6 per cent high, itself carrying a 5 per cent standard deviation.

Each true mass is transformed by the measurement chain before it enters the published equations. Writing the chain out makes the propagation reproducible and shows which terms are proportional and which are additive:

$$m = m_{\text{true}}(1+\delta)(1+\varepsilon)(1+w) + m_{\text{true}}\tau \tag{9}$$

where

m = mass as recorded, in kilograms
m_{true} = true mass of the consignment
δ = proportional calibration drift of the instrument, drawn once per instrument
ε = proportional weighing noise, normal with mean zero and standard deviation 0.004
w = proportional moisture and debris term, zero unless the consignment is weighed wet
τ = standing tare offset of the instrument, zero unless the tare is mis-set

The drift term accumulates with the time since the instrument was last calibrated, which is what makes the calibration interval bite:

$$\delta = \sigma \times g \times \frac{t}{30} \tag{10}$$

where

σ = direction of drift, plus or minus one, a property of the instrument
g = drift rate, 0.0008 proportional bias per thirty days
t = days since last calibration, drawn uniformly over the interval the regime permits

These parameter values are assumptions, not measurements, and every one of them is stated above so that a reader may disagree with any of them and re-run the analysis. The reporting here is deliberately not in the form of the Guide to the Expression of Uncertainty in Measurement [11]: no combined standard uncertainty, coverage factor or expanded uncertainty is computed, because the inputs are assumed rather than evaluated from measurement data and a GUM budget built on assumed inputs would carry an authority it has not earned. The Guide is cited here only for the distinction between effects that are corrected and effects that are left in the result, which is the distinction the dry-weight and calibration controls turn on. The results that follow are differences between configurations under common assumptions, which is the comparison the assumptions can support. Absolute bias magnitudes depend on the parameter values and should not be read as estimates of error in any real project.

Two factors are crossed: the calibration regime, at four levels, and the presence or absence of a dry-weight rule, giving eight cells. The cells are not simulated independently: replication r of every cell is run from the same seeded stream, so the eight cells see the same sequence of true masses and the same sequence of error draws, and differ only in how the configuration transforms them. This is the method of common random numbers, and it is used deliberately, because the quantity of interest is the difference between configurations and pairing removes most of the sampling noise from that difference. The ideal regime draws its drift time and direction like every other regime and then discards them, so that it stays aligned with the others on the same stream rather than running two draws ahead. Each programme is then simulated at the cell matching the provisions actually located in its documents: Verra at three-yearly calibration with a dry-weight rule, PCX at proof-at-issuance calibration without one, and OBP and CCM at no calibration without one.

Each replication is scored as the bias of the reported net quantity against the true net, and the reported figures are the mean and standard deviation of that quantity over replications:

$$\beta_r = \frac{N_{\text{reported,r}} - N_{\text{true,r}}}{N_{\text{true,r}}} \tag{11}$$

$$\text{SE}_{\text{MC}} = \frac{s}{\sqrt{R}} \tag{12}$$

$$\text{SE}_{\text{paired}} = \frac{s_d}{\sqrt{R}} \tag{13}$$

where

β_r = bias on replication r, reported as a percentage
N = net credited quantity from Equation 4 of the applicable methodology

s = standard deviation of the per-replication bias

s_d = standard deviation of the per-replication difference between two configurations run on the same stream

R = number of replications, here 2,000

The second of these is the Monte Carlo standard error of a reported mean and the third the standard error of a paired contrast under common random numbers; a difference smaller than a few of the relevant error is simulation noise and is reported as such. Amplification through the net-quantity equation is the ratio

$$A = \frac{\beta_{\text{net}}}{\beta_{\text{gross}}} \tag{14}$$

where

β_{net} = bias on the net quantity, project minus baseline

β_{gross} = bias on the gross project quantity alone

VI. RESULTS

6.1 The two controls act independently

The two principal controls act on different properties of the error distribution, and they act almost independently. Figure 1 shows the separation.

A dry-weight rule reduces mean bias in the credited mass from 0.85 per cent to 0.22 per cent, a reduction of roughly 3.8-fold, while leaving the spread of that bias almost unchanged at 1.183 against 1.198 per cent. The calibration regime does the opposite. Across all four regimes mean bias is effectively constant, moving only from 0.548 per cent under ideal calibration to 0.514 per cent under none. The reason is that an uncalibrated scale is as likely to read high as low: within any one project its drift is a standing bias, but across projects the two directions are equally likely, so the average over replications is zero however long the interval. What the calibration interval changes is the standard deviation, which widens from 0.274 per cent under ideal calibration to 3.211 per cent with no calibration rule, a factor of about 11.7.

Both contrasts are paired: every configuration is run on the same seeded random stream for each replicate, so the difference between two configurations is computed replicate by replicate and its Monte Carlo error is the error of those differences rather than of either mean separately. The dry-weight contrast is -0.623 percentage points with a paired standard error of 0.0019, a difference of some three hundred standard errors, which is not in doubt. The contrast between no calibration rule and perfect calibration is -0.0337 percentage points with a paired standard error of 0.0719: smaller than its own error, and therefore no effect on the mean at all. The calibration regime should be read as moving dispersion and nothing else, and no ordering of mean bias across calibration regimes in Table 3 is a finding.

One caveat attaches to the “ideal” regime wherever it appears below. It is not a calibration practice that any programme requires or that any operator could achieve; it is a benchmark in which calibration drift is set to zero, so that the drift contribution can be read off by difference. Its standard deviation of 0.274 per cent is a lower bound on achievable dispersion under the remaining error sources, not an attainable target, and it should not be quoted as what good calibration delivers. The attainable comparison is between proof-at-issuance and no rule, where the standard deviation widens from 0.284 to 3.211 per cent.

This separation is the study's central result for control design, and it is not obvious in advance. A programme that regulates calibration thoroughly and says nothing about moisture has bought precision without accuracy: its credited quantities will be tightly clustered around a value that is systematically too high. A programme that requires dry weight but leaves calibration unregulated has the opposite profile, centred correctly but widely dispersed. Neither control substitutes for the other, and a comparison of programmes on whether they “have measurement requirements” will not distinguish them.

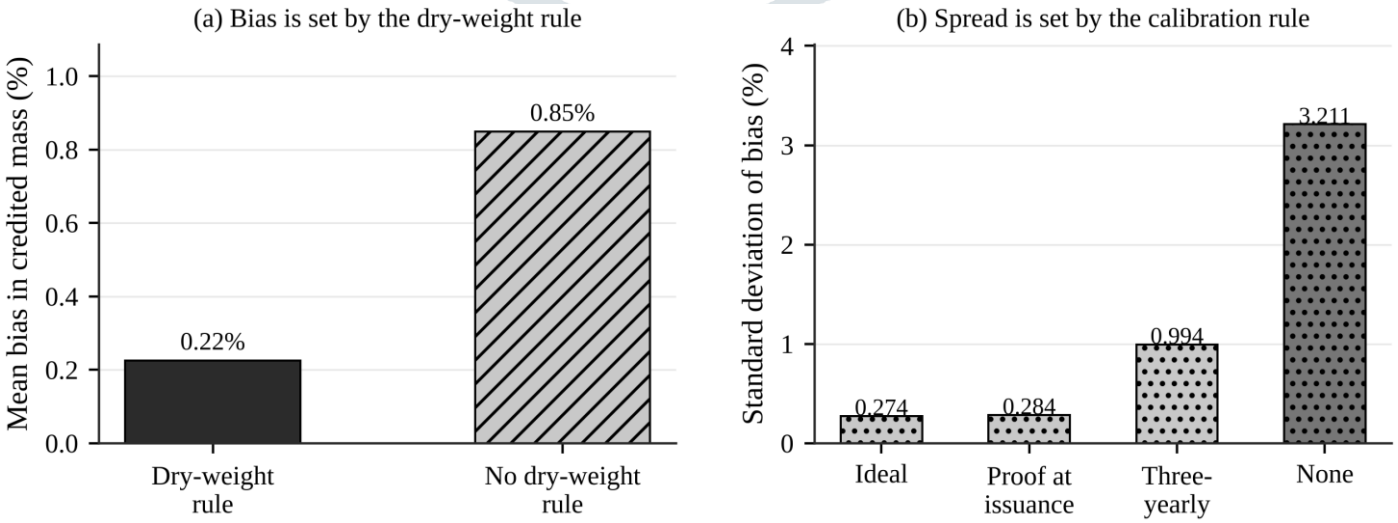

Figure 1: (a) mean bias in credited mass with and without a dry-weight rule; (b) standard deviation of that bias under four calibration regimes

6.2 Each programme at its own configuration

Table 3 and Figure 2 report each programme simulated at its own configuration. Verra, with three-yearly calibration and a dry-weight rule, shows an expected bias of 0.23 per cent with a 95 per cent interval of -1.64 to 2.12 per cent. PCX, with the strongest calibration requirement but no moisture deduction, shows the tightest interval of any programme, 0.19 to 1.60 per cent, around a mean of 0.86 per cent. OBP and CCM, with neither control, share an expected bias of 0.83 per cent and the widest interval, -5.35 to 6.95 per cent.

The PCX result repays attention because it is counter-intuitive. PCX operates the most demanding calibration control in the sector, requiring proof of calibration to be uploaded with every issuance rather than merely requiring the practice. That control works: PCX has the narrowest distribution of any programme modelled, narrower than Verra's. But its expected bias is about 3.8 times Verra's, because calibration does not address what is being weighed. A programme can hold the best measurement-equipment rule in the sector and still systematically over-issue.

Two comparisons in Table 3 must not be drawn. OBP and CCM are modelled at the same configuration, because neither publishes a calibration rule or a dry-weight rule, so their rows are identical by construction and carry no information about any difference between the two programmes. And the separation between the PCX and OBP means, 0.859 against 0.826 per cent, is smaller than the Monte Carlo error on the contrast and is simulation noise; what distinguishes them is dispersion, where PCX's standard deviation of 0.291 per cent against OBP's 3.222 per cent is an order-of-magnitude difference that the error cannot explain.

All four biases are positive. Under these assumptions every error source that is one-directional runs the same way: wet material reads high, adherent debris reads high, and estimates substituted for missing measurements run high. No programme applies an uncertainty discount or conservativeness factor that would offset this, and none states a mass-balance tolerance that would bound it. The magnitudes here are small, below 1 per cent, and follow from the chosen parameters; the direction does not depend on the parameters and follows from the structure of the error sources.

Table 3: Expected bias in credited mass under each programme's configuration of controls

Programme	Calibration rule	Dry weight	Mean bias %	MC SE %	95% interval %	SD %
Verra	At least every three years	Yes	0.23	0.0221	-1.64 to 2.12	0.989
PCX	Proof uploaded at issuance	No	0.86	0.0065	0.19 to 1.60	0.291
OBP	No provision located	No	0.83	0.0720	-5.35 to 6.95	3.222
CCM	No provision located	No	0.83	0.0720	-5.35 to 6.95	3.222

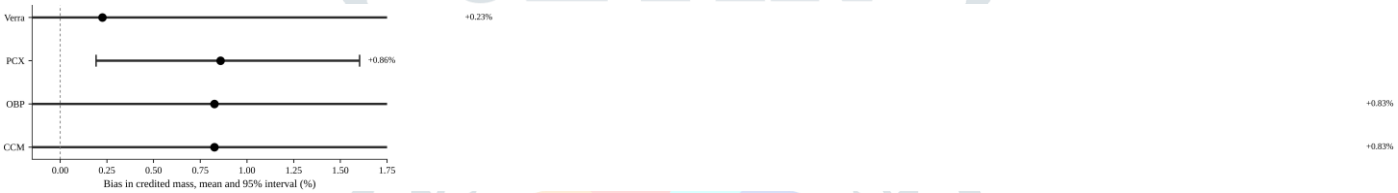

Figure 2: mean bias and 95 per cent interval in credited mass, by programme

6.3 Amplification of bias through the net-quantity equation

A third result comes from running the error through PWRM0001 Equation 4 rather than through a single mass, and it turns on how the baseline is established. Where baseline and project quantities are weighed in the same period under the same regime, a proportional bias applies equally to both and passes through the difference unchanged: gross bias of 0.85 per cent gives net bias of 0.86 per cent, an amplification of 1.0. Where the baseline instead comes from historical records not produced under the project's current measurement regime, only the project side carries the bias, and a gross bias of 0.77 per cent becomes 1.23 per cent on the net, an amplification of 1.59. The two gross figures are close but not identical, 0.85 against 0.77 per cent, because dropping the baseline weighings also removes their draws from the random stream; the comparison that matters is between each run's own gross and net. This is not a hypothetical configuration: PWRM0001 sets a capacity-addition baseline from the maximum collection potential of the three years before the project started, which is by construction a historical figure. The amplification is a property of the published equation, not of the error model.

6.4 Baseline convention as a source of methodological uncertainty

A fourth result concerns methodological rather than measurement uncertainty, and it is internal to Verra. The two methodologies set a capacity-addition baseline differently. PWRM0001 determines it "based on the maximum collection potential of the waste collection system in the three-year period prior to the start of the project activity". PWRM0002 uses the average annual recycling rate over the three prior years. Maximum and average are not interchangeable: for any series with variation, the maximum exceeds the mean, so a maximum-based baseline is the more conservative convention and yields a smaller credited net.

Simulating 3,000 projects with a base level drawn uniformly between 800 and 5,000 tonnes a year, three prior years varying independently within plus or minus 20 per cent of that base level, and a project quantity between 1.05 and 1.60 times the prior maximum — all three ranges being the author's, since no programme publishes a distribution of prior-year variability — the average-based convention yields a credited net 41.1 per cent higher at the mean and 27.0 per cent higher at the median than the maximum-based convention. The fifth to ninety-fifth percentile range of the difference runs from 6.8 to 129.0 per cent. The dispersion is wide because the gap between a maximum and a mean depends on how variable the prior years were, and it grows as the project increment shrinks relative to that variation.

The comparison is between two conventions applied to the same project structure, not between collection and recycling activities, which differ in other ways too. What it isolates is the effect of the convention itself. A difference of this size, arising within a single programme from a choice of word in two documents, is larger than any measurement effect examined in this paper, and it is not disclosed to a credit buyer anywhere in the credit record.

6.5 What the records can and cannot reveal

A control framework is only as useful as the errors it can actually see. Table 4 classifies each error source by whether it can be detected from the transaction records alone. The classification is reasoned from the structure of each error and from the control rules examined in the companion chain-of-custody study; it is not an output of the simulation reported above, which models the weighing operation rather than a record-level detector, and three of the rows concern errors the simulation does not generate.

The errors that records catch are the discrete ones: a unit slip, a broken mass balance, an inverted timestamp, a missing event. These leave a trace in the data because they make the record internally inconsistent, and automated controls find them reliably.

The errors that records cannot catch are precisely the ones that drive bias. A scale reading 0.8 per cent high produces a figure that is internally consistent with every other figure in the chain; nothing in the records contradicts it, and only the calibration certificate reveals it. Wet material weighed as though dry produces a valid mass for the material as presented; detecting it requires either a physical inspection at the weighbridge or a procedural rule that the material be dried and cleaned first. The baseline convention effect is not an error in the records at all, since the arithmetic is correct under either convention.

This has a direct consequence for how assurance effort should be allocated. Document review and automated reconciliation, which is where most verification effort goes, addresses the variance-type errors and the discrete recording failures. The bias-type errors that accumulate in the credited total are reachable only by calibration evidence and site inspection. A verification regime weighted towards document review will therefore detect the errors that largely cancel and miss the errors that compound, which is the opposite of the desirable allocation.

Table 4: Detectability of each error source from transaction records, and the residual need for human review

Error source	Detectable from records alone	What detects it	Residual need for human review
Unit transcription error	Yes, reliably	Plausibility bound on mass for the declared unit	None; the error is three orders of magnitude
Gross mass-balance break	Yes	Retention outside the declared band for that transition and material	Attribution: which of the two records is wrong
Timestamp inversion	Yes	Ordering check across the chain	None
Missing measurement	Yes, if estimates are flagged	Completeness check against expected events	Whether the estimate is reasonable
Stale or wrong tare weight	Only if it breaks a band	Mass reconciliation, indirectly	Yes; a small tare error is within tolerance and invisible
Scale calibration drift	No	Nothing in the record; the figure is internally consistent	Yes; requires the calibration certificate, not the record
Moisture and debris in weighed mass	No	Nothing in the record; wet mass is a valid number	Yes; requires physical inspection or a dry-weight procedure
Undocumented correction	Only as an absence	Missing source or correction reference	Yes; unresolvable from records, by definition
Baseline convention effect	No	Nothing; the arithmetic is correct under either convention	Yes; a methodology question, not a records question

VII. DISCUSSION

Four implications follow.

The first is that measurement requirements should be specified against the error they address rather than as a general obligation to measure accurately. Bias and variance are controlled by different rules, and a programme can score well on one while leaving the other untouched. The practical test for a standard setter is not whether a measurement requirement exists but whether, for each one-directional error source, some provision removes or bounds it. On the evidence here, three of four programmes have no provision addressing the largest one-directional source in the chain.

The second concerns the gap between conservativeness as a principle and conservativeness as a number. Verra and PCX both require a conservative approach, and Verra's formulation is binding and specific about direction: net quantities must not be overestimated. But a principle binds a judgement, not an arithmetic operation, and no programme in the corpus attaches a numeric deduction or buffer to the credited figure. Because the one-directional error sources all run upward, the expected effect of leaving uncertainty unadjusted is systematic over-issuance rather than a symmetric distribution of error about the true value. The magnitude under these assumptions is small; the direction is not an artefact of the assumptions.

The third concerns what a verifier can act on. Only PCX requires that a discrepancy in the chain of custody be analysed and explained, and it states no permitted variance, which leaves the verifier to decide when a difference between two weighings is material. The other three state neither the duty nor the tolerance: Verra publishes no reconciliation threshold, OBP requires a mass balance without one, and the CCM monitoring template carries input, output and loss totals with no stated relationship between them. Without a stated tolerance, two verifiers applying the same standard to the same records may reasonably differ, and the absence of the threshold is therefore a reproducibility problem as much as a rigour one. A tolerance band also has to be material-specific, and the programmes' own documents supply the reason: PWRM0002 assigns a default plastic fraction of 1.00 to a non-composite resin and 0.04 to an unspecified composite, a twenty-five-fold difference in how much of a weighed input the methodology expects to survive as plastic. A reconciliation tolerance that ignores that difference will either mask real discrepancies in the high-yield streams or raise constant false alarms in the low-yield ones. The paper makes no claim about actual process yields, which are not published by any programme in the corpus.

The fourth concerns documentation quality in the methodologies themselves. Every mathematical-italic glyph in both published PDFs is corrupted on text extraction, so none of the eight equations could be copied reliably out of the document as published; six were recovered only by rendering the page as an image and reading the typeset formula, and the remaining two were accepted from extracted text because they are simple summations whose definition lists survived intact. One of the recovered equations, the penetration-rate formula, uses identical notation for different quantities in two methodologies of the same programme. A third carries a notation inconsistency between its equation and its own monitoring table. None of these is a flaw in the underlying method, and all are recoverable by a reader who renders the pages. But a quantification rule that cannot be copied accurately out of its own publication is a barrier to the independent scrutiny that voluntary standards depend on, and the remedy, publishing equations in an accessible text form alongside the typeset version, is inexpensive.

VIII. A MEASUREMENT CONTROL CHECKLIST

The following controls are ordered by the error they address and by what the simulation shows each to be worth.

1. Weigh on a dry-weight basis, net of non-plastic substances, and record the basis used. This is the only control examined that reduces systematic bias, removing 0.62 percentage points of it on a paired comparison (Section 6.1), which is a reduction of about 3.8-fold against the same configuration without the rule. Three of the four programmes do not require it.

2. Calibrate weighing equipment on a stated interval and retain the certificate. Calibration does not reduce bias — the paired contrast on the mean is within its own Monte Carlo error — but it controls spread, narrowing the standard deviation by a factor of about 11.3 between no rule and a proof-at-issuance rule. Requiring the proof at the point of claim, as PCX does, is materially stronger than requiring the practice and leaving it to audit.

3. Record the unit of measure explicitly on every document and apply a plausibility check. This matters more across programmes than within one: an OBP Credit represents one kilogram of non-commercially-recyclable ocean bound plastic, while Verra, PCX and CCM all denominate in metric tonnes. A thousand-fold difference in the credit unit between programmes sits alongside a thousand-fold transcription error as the same arithmetic mistake, and anyone aggregating across programmes has to carry the distinction explicitly. A plausibility check that compares each recorded mass against the expected range for its container or vehicle will catch a unit transcription before it reaches a credit. Unit errors are rare but three orders of magnitude in size, which makes them the highest-consequence recording error and also the easiest to detect automatically.

4. Re-measure tare weights on a stated schedule rather than carrying a standing figure. A stale tare is a silent error that no downstream reconciliation will attribute correctly.

5. Flag every estimated value as an estimate and reconcile estimates against measurement where measurement later becomes available. Estimates substituted for missing measurements run high and are otherwise indistinguishable from measured values in the record.

6. Set a material-specific reconciliation tolerance, publish it, and treat departures from it as findings. Only PCX requires a discrepancy to be analysed at all, and no programme states a number, so at present a verifier deciding whether a given loss is acceptable is deciding it unaided; publishing the band turns that judgement into a check that two verifiers would answer the same way.

7. Record a correction reference whenever a recorded value changes. An undocumented correction produces uncertainty that cannot be quantified at all, which is a worse position than a known error of known size.

8. State which baseline convention applies and disclose it with the credit. Within a single programme, the choice between a maximum-based and an average-based baseline moved the credited net by 41 per cent at the mean in simulation, which exceeds every measurement effect examined here.

IX. LIMITATIONS

The simulation propagates stated assumptions through published equations. Every parameter governing error magnitude and frequency was chosen by the author, and none was estimated from operating data, because such data are not published. The absolute bias figures are therefore properties of the assumption set. What the assumptions can support is the comparison between configurations, since all configurations share them, and the qualitative findings on which control acts on bias and which on variance, which follow from the structure of the error sources rather than from their magnitudes.

The simulation models the weighing operation and its propagation into the credited total. It does not model the full multi-stage mass balance across sorting fractions, rejected material and inventory movement; a companion study by the author, in preparation, of chain-of-custody evidence simulates that structure separately, with a seven-stage chain and a published process model. The two are complementary: the present study asks how accurate each weighing is, the companion study asks whether successive weighings reconcile. Unreconciled stock movement in particular is outside the scope modelled here.

Error sources are modelled as independent of one another, with one exception: calibration drift and a mis-set tare are properties of an instrument and are held fixed across every weighing that instrument performs, so errors from those two sources are correlated within a project by construction. The others are not. In a real facility they would be: a facility that neglects calibration is plausibly the same facility that neglects tare re-measurement and substitutes estimates for missing weighings. Correlated failures would widen the distributions beyond those reported here, so the figures for the unregulated configurations should be read as conservative.

The mapping from programme to configuration is a reading of each programme's documents, and it is binary where reality is graded. A programme with no calibration provision may still have projects that calibrate, because commercial weighbridges are often calibrated for trade reasons regardless of what a credit standard requires. The simulation reports what each programme's rules guarantee, not what its projects do.

The absence claims for OBP and CCM calibration rest on full reading of the operative instruments identified for each programme, and are stated as no provision located rather than as certainty of absence. A provision in a document outside the corpus examined would change the configuration and therefore the result.

The evidence base is not of uniform depth across the four programmes, and the PCX findings rest on the narrowest of the four. Verra, OBP and CCM are each represented by a layered set of instruments — standard, methodology or component standards, auditing protocol and templates — whereas PCX publishes a single five-module standard with no accompanying quantification methodology, so there is no PCX document in which a quantification equation, a moisture deduction or a reconciliation tolerance could have appeared. Statements here that PCX does not quantify in a particular way therefore compare a present provision in one programme with an absent document layer in another, and should be read as findings about what the published standard does not address rather than as evidence of a weaker rule. The PCX calibration and chain-of-custody provisions relied on, by contrast, are in the standard itself and are primary evidence of the same class as the others.

The baseline asymmetry simulation compares two conventions on a synthetic project structure with prior-year variation drawn uniformly within plus or minus 20 per cent. Real collection and recycling series would have different variability, and the size of the difference between a maximum and a mean depends directly on that variability. The direction of the difference does not.

Finally, this study examines quantification rules and their propagation. It does not examine whether projects comply with those rules, which would require facility records, nor whether verifiers detect non-compliance, which would require validation reports.

X. CONCLUSION

This study compared measurement and reconciliation requirements across four plastic credit programmes, verified the eight quantification equations that one of them publishes, and propagated measurement error through those equations under each programme's configuration of controls.

Two of the four programmes, Ocean Bound Plastic and the Circular Credits Mechanism, contain no weighing-equipment calibration provision of any kind. Three of the four contain no moisture or contamination deduction; only Verra requires weighing on a dry-weight basis net of non-plastic substances. None of the four applies an uncertainty discount or states a mass-balance reconciliation tolerance.

The simulation shows the two principal controls to be close to orthogonal. A dry-weight rule cuts mean bias from 0.85 to 0.22 per cent and barely touches spread; the calibration interval leaves mean bias unchanged, the paired contrast falling within its own Monte Carlo error, and widens the standard deviation from 0.274 to 3.211 per cent, the lower figure being a zero-drift benchmark rather than any achievable practice. Configured as each programme specifies, expected bias is 0.23 per cent for Verra, 0.86 per cent for PCX and 0.83 per cent for OBP and CCM. PCX holds the sector's strongest calibration requirement and the narrowest distribution, yet several times Verra's expected bias, because calibration governs how well material is weighed and not what is weighed.

A separate finding is internal to Verra. Its collection methodology sets a capacity-addition baseline on the maximum of three prior years and its recycling methodology on the average. Simulation puts the resulting difference in credited net at 41 per cent at the mean, larger than any measurement effect examined here and not disclosed with the credit.

The simulation protocol, its parameters and the measurement-control checklist are published so that each can be revised and re-run. Validating the assumptions requires weighbridge records, calibration certificates and facility input-output data, and obtaining them is the necessary next step.

DECLARATIONS

Declaration on the use of artificial intelligence [12]. In preparing this manuscript the author used a generative artificial intelligence assistant (Anthropic Claude) to support the structured coding of programme documents, the generation and analysis of the synthetic datasets described in the methodology, the drafting of text and the formatting of the manuscript. The research design, the selection of source documents, the interpretation of the coded results and all substantive conclusions are the author's own. The author reviewed and verified the content of the manuscript, including every quantitative result and every cited source, and takes full responsibility for its accuracy and integrity. No artificial intelligence tool is listed as an author, in accordance with COPE guidance on authorship and artificial intelligence [12].

Declaration of competing interests. The author declares no financial or non-financial competing interests in relation to the plastic credit programmes examined in this study. The author is not affiliated with, funded by, or contracted to Verra, the Ocean Bound Plastic Certification Programme, PCX Solutions or the Circular Credits Mechanism.

Funding. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Ethics statement. This study analyses publicly available programme documentation and computer-generated synthetic records only. It involves no human participants, no personal data and no primary data collection, and therefore did not require ethical approval.

Data availability. All programme documents analysed are publicly available from the respective programme websites cited in the references. The coding protocol and every parameter value governing the reported results are stated in the methodology. The procedure is stochastic and is seeded with 20261030, so rerunning the author's script reproduces the reported figures exactly. An independent reimplementation will reproduce them in expectation rather than digit for digit, because the figures depend on the generator and on the order in which it is drawn from. No data repository accompanies this article; the author's own scripts are available on request.

ACKNOWLEDGMENT

The author thanks Verra for publishing its quantification methodologies in full, which made the equation-level analysis in this paper possible, and Zero Plastic Oceans, PCX Solutions and BVRio for the public availability of their standards.

REFERENCES

[1] Verra. 2021. Plastic Waste Reduction Standard, Version 1.0. Washington DC: Verra.
[2] Verra. 2022. PWRM0001 Plastic Waste Collection Methodology, Version 1.1. Washington DC: Verra.
[3] Verra. 2022. PWRM0002 Plastic Waste Recycling Methodology, Version 1.1. Washington DC: Verra.
[4] Zero Plastic Oceans. 2026. OBP Collection Organization Standard, OBP-COL-STD Version 2.3, release date 8 September 2026. Zero Plastic Oceans.
[5] Zero Plastic Oceans. 2026. OBP Recycling Organization Standard, OBP-REC-STD Version 2.4, release date 8 September 2026. Zero Plastic Oceans.
[6] Zero Plastic Oceans. 2026. OBP Program Definitions and Annexes, OBP-DEF-GUI Version 1.6, release date 8 September 2026. Zero Plastic Oceans.
[7] PCX Solutions. 2024. Plastic Pollution Reduction Standard, Version 8.0, effective 5 June 2024. Manila: PCX Solutions.
[8] BVRio. 2024. Circular Credits Mechanism Standard, Version 2.0, November 2024. BVRio Institute.
[9] BVRio. 2022. Circular Credits Mechanism Auditing Protocol, Version 2.2. BVRio Institute.
[10] BVRio. 2022. Circular Credits Mechanism Monitoring Report Template, Version 1.4. BVRio Institute.
[11] Joint Committee for Guides in Metrology. 2008. JCGM 100:2008, Evaluation of measurement data - Guide to the expression of uncertainty in measurement. Sevres: BIPM.
[12] Committee on Publication Ethics. 2023. Authorship and AI tools: COPE position statement. Committee on Publication Ethics.